

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 0 813 158 A3**

(12) **EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:  
21.04.1999 Bulletin 1999/16

(51) Int. Cl.<sup>6</sup>: **G06F 17/30**

(43) Date of publication A2:  
17.12.1997 Bulletin 1997/51

(21) Application number: **97109680.5**

(22) Date of filing: **13.06.1997**

(84) Designated Contracting States:  
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**

(30) Priority: **14.06.1996 US 661335**

(71) Applicant: **Excite, Inc.**  
**Mountain View, California 94043 (US)**

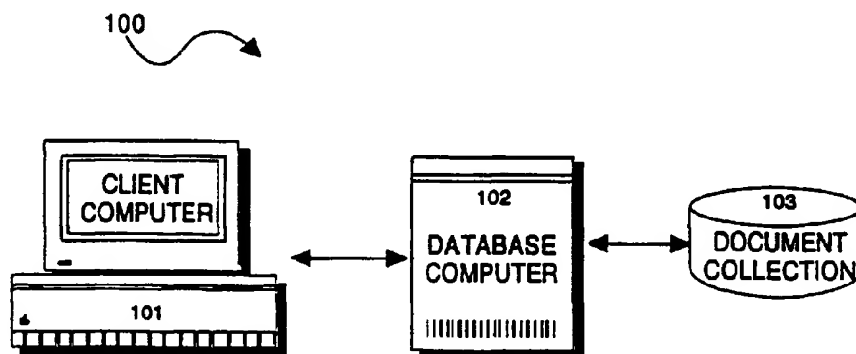
(72) Inventor: **Spencer, Graham**  
**Cupertino, California 95014 (US)**

(74) Representative: **Liesegang, Eva**  
**Forrester & Boehmert,**  
**Franz-Josef-Strasse 38**  
**80801 München (DE)**

(54) **System and method for accelerated query evaluation of very large full-text databases**

(57) A system, method, and various software products provide for improved information retrieval in very large document databases through the use of a predetermined static cache. The static cache includes for terms that appear in a large number of documents, a plurality of documents ordered by a contribution that the term makes to the document score of the document. The contribution is a scalar measure of the influence of the term in the computed document score. The contribution reflects both the within document frequency and the between document frequency of the term. In addition, the static cache includes for each term a lookup

table that references selected entries for the term in an inverted index. Queries to the database are then processed by first traversing the static cache and obtaining the contribution information therefrom and computing the document score from this information. Additional term frequency information for other terms in the query is obtained by looking up the document in the lookup tables of the other query terms, and obtaining the term frequency information for such terms from the inverted index, or by searching the contribution caches of the query terms.



**FIGURE 1**

**EP 0 813 158 A3**



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 97 10 9680

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	PERSIN M: "DOCUMENT FILTERING FOR FAST RANKING" SIGIR '94, DUBLIN, JULY 3 - 6, 1994, no. CONF. 17, 3 July 1994, pages 339-348, XP000475332 CROFT W B; VAN RIJSBERGEN C J (EDITORS) * abstract * * paragraph 2.2 * * paragraph 3.2-3.3 * ---	1,3,5-31	G06F17/30
A	AALBERSBERG I J: "A DOCUMENT RETRIEVAL MODEL BASED ON TERM FREQUENCY RANKS" SIGIR '94, DUBLIN, JULY 3 - 6, 1994, no. CONF. 17, 3 July 1994, pages 163-172, XP000475321 CROFT W B; VAN RIJSBERGEN C J (EDITORS) * the whole document * ---	1-31	
A	WONG W Y P ET AL: "IMPLEMENTATIONS OF PARTIAL DOCUMENT RANKING USING INVERTED FILES" INFORMATION PROCESSING & MANAGEMENT (INCORPORATING INFORMATION TECHNOLOGY), vol. 29, no. 5, September 1993, pages 647-669, XP002035616 * the whole document * ---	1-31	TECHNICAL FIELDS SEARCHED (Int.Cl.6) G06F
A	US 5 263 159 A (MITSUI KINICHI) 16 November 1993 * abstract * * column 1, line 1 - column 2, line 47 * -----	1-31	
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 26 February 1999	Examiner Katerbau, R
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03/82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT  
ON EUROPEAN PATENT APPLICATION NO.**

EP 97 10 9680

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

26-02-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5263159 A	16-11-1993	JP 1945979 C	23-06-1995
		JP 3108064 A	08-05-1991
		JP 6075265 B	21-09-1994
		DE 69031772 D	15-01-1998
		DE 69031772 T	25-06-1998
		EP 0420424 A	03-04-1991
-----			

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82